PROCEEDINGS OF SCIENCE

Approximation of high-resolution surface wind speed in the North Atlantic using discriminative and generative neural models based on RAS-NAAD 40-year hindcast

V. Y. Rezvov,[a,b,*] M. A. Krinitskiy[a,b] and S.K.Gulev[a]

[a] *Shirshov Institute of Oceanology, Russian Academy of Sciences, 36, Nahimovskiy prospekt, Moscow, Russia*

[b] *Moscow Institute of Physics and Technology, 9, Institutskiy per., Dolgoprudny, Moscow Region, Russia*

E-mail: rezvov.vyu@phystech.edu

Surface wind is one of the most important atmospheric fields in climate research. Accurate prediction of high spatial resolution surface wind has a wide variety of applications, such as renewable wind energy and forecasts of extreme weather events. General circulation models (GCMs) study climate system on a global scale. Their main issues are the low resolution of the modeling results and high computational costs. One of the solutions to these problems is statistical downscaling. Statistical downscaling methods discover functional relationships avoiding computationally expensive high-resolution hydrodynamic simulations. Deep learning methods, including artificial neural networks (ANNs), are one of the typical machine-learning approaches approximating complex nonlinear relationships. In our study, we explored the capabilities of statistical 5x spatial downscaling of surface wind over the ocean in the North Atlantic region. Low-resolution input data and high-resolution validation data were provided by RAS-NAAD 40-year hindcast. We applied several downscaling methods, including bicubic interpolation as a reference solution, various discriminative convolutional neural networks (CNNs) such as Linear CNN, Residual CNN, CNN with skip connections, and generative adversarial network (GAN) based on SR-GAN. We also compared downscaling results in terms of RMSE, PSNR and other quality metrics including the ones representing the reconstruction of extreme winds. We evaluated the computational costs and the quality of different methods and reference solution to identify advantages and lacks of machine-learning downscaling. As a result, both discriminative and generative ANN-based downscaling methods have not outperformed reference solution in downscaling quality. Nevertheless, for further research, we consider GANs as the most promising ANN architectures for surface wind downscaling based on their fine-structure modeling ability.

The 6th International Workshop on Deep Learning in Computational Physics (DLCP2022)
6-8 July 2022
JINR, Dubna, Russia

*Speaker

PoS(DLCP2022)023

1 Introduction

Climate change is one of the most serious problems of the modern world. Because of global warming, temperature rises and changes local precipitation patterns. In particular, these phenomena are explained by local changes of wind speed [1].

General circulation models (GCMs) are mathematical models describing physical processes and interactions in the Earth-atmosphere-ocean system. They study climate system on a global scale. The main problem of GCMs is the low resolution of the modeling results. In particular, the computational cells of GCMs are too large compared to the spatial size of the weather phenomena to be investigated. The low resolution of GCM outputs causes systematic errors and implausible future climate scenarios, especially for extreme weather events [2]. In addition, GCMs, including low-resolution numerical models, are computationally expensive [3].

One of the solutions to this problem is downscaling. This approach obtains high-resolution information about physical variables from low-resolution modeling outputs. The downscaling methods consist of two distinct groups: dynamical downscaling and statistical downscaling [4].

Dynamical downscaling applies both low-resolution and high-resolution numerical modeling. A coarser model calculates outputs in an entire modeling area. Then, these outputs are used as boundary conditions for a high-resolution model in particular subareas of the modeling area [5-6]. This approach significantly reduces computational costs because it does not simultaneously provide high-resolution modeling in all parts of the region. However, the performance of dynamical downscaling methods is still insufficient.

Statistical downscaling techniques avoid high-resolution numerical simulation. In this group of methods, the functional relationship between low- and high-resolution data is approximated by training a statistical model on a special dataset. The quality of statistical downscaling is comparable to that obtained in dynamical downscaling [7–8]. Nevertheless, in practice, statistical methods are widely used because they have low computational costs.

To date, a large number of statistical downscaling methods have been developed. The research has shown that the most promising group of statistical methods was statistical regression [9–10]. Its examples are multilinear regression [11–13], generalized linear models [14], quantile regression analysis [15].

Besides classical statistical approaches, statistical downscaling also includes methods that are more efficient and fast [16]. Some publications have discussed efficiency of nonlinear regression methods [17–18]. Recent considerable scientific interest has been generated by contemporary machine-learning techniques for statistical downscaling. For example, the Relevance Vector Machine (RVM) [19] and various types of artificial neural networks (ANNs), such as autoencoders [18], recurrent neural networks [20] and convolutional neural networks (CNNs) [1, 21], have been investigated.

A problem of statistical downscaling of the surface wind was solved in [22–24]. A comparative analysis of various statistical downscaling methods for climatic variables was carried out in [18, 21, 25–26]. The authors of [18] have found that non-classical machine-learning approaches did not result in a better quality of climatic downscaling. The authors of [25] have concluded that machine-learning methods provided better downscaling of wind. In [21], convolutional neural networks were compared with the classical linear model. The study results have shown that CNNs downscaled variables more accurately. The authors of [26] have trained a

fully-connected artificial neural network (FCNN) to downscale wind speed in North America. They showed that the FCNN-based model had outperformed classical linear regression.

Summing up, recent advances in artificial neural networks and the emergence of contemporary machine-learning techniques have led to a significant increase in the popularity of statistical downscaling. However, the insufficient number of publications does not allow to determine the quality of non-linear machine learning methods compared to classical ones. In addition, the results of the research on this topic are highly contradictory, and several questions remain relevant. Therefore, assessment of contemporary methods of statistical downscaling for the climate is still a task of current scientific interest.

2 Initial data

We used a retrospective NAAD dynamic model developed by Shirshov Institute of Oceanology (Russian Academy of Sciences) in collaboration with the Institut des Géosciences de l'Environnement. The model outputs were the atmospheric fields in the North Atlantic region [27].

NAAD uses the non-hydrostatic WRF 3.8.1 model [28]. The modeling area covers the North Atlantic region from 10º N to 80ºN and from 90ºW to 5ºE. The center of the area is at the point with the coordinates (45ºN, 45ºW).

The initial and boundary conditions for the NAAD model, including the sea surface temperature, were taken from the ERA-Interim reanalysis [29]. In the HiRes experiment of the NAAD, the modeling area was a regular grid of 110×110 nodes. The distance between the nodes was approximately 14 km, and the lower level was 10–12 m above the ocean surface. The NAAD model also performed the LoRes experiment on a regular grid of 550×550 nodes, and the distance between them was 77 km. All NAAD experiments were carried out over a 40-year period from January 1979 to December 2018.

We used the LoRes dataset of the NAAD model with a three-hour time resolution as the low-resolution input data for the statistical downscaling models. High-resolution target dataset was the NAAD HiRes dataset with the same time intervals. Our analysis was limited to a 38-year time period from January 1979 to December 2016. Both the low-resolution input variables and the high-resolution target variables consisted of two near-surface wind speed orthogonal horizontal components and the sea-level atmospheric pressure.

We chose the bicubic interpolation as a reference solution to compare with artificial neural networks.

3 Discriminative artificial neural networks

In discriminative artificial neural networks used for downscaling, the low-resolution values of physical variables are outputs of rough numerical simulation. These results are an argument x for a certain function f. The value of the function $z^* = f(x)$ should be, ideally, equal to a true high-resolution value z of the desired physical variable. The functional relationship f between low- and high-resolution data is approximated by training a discriminative ANN on a special dataset with low- and high-resolution data pairs.

As an example of discriminative ANNs we used convolutional neural networks (CNNs) for wind downscaling. The CNNs are parametric mappings that optimize specific model parameters

during the model training. This approach identifies various abstract features of the input data that is important for climatic dataset. In CNNs, a fixed-size convolution kernel is sequentially applied to regular input data. For each convolutional layer, the convolution kernel is a set of trained parameters in the form of an array.

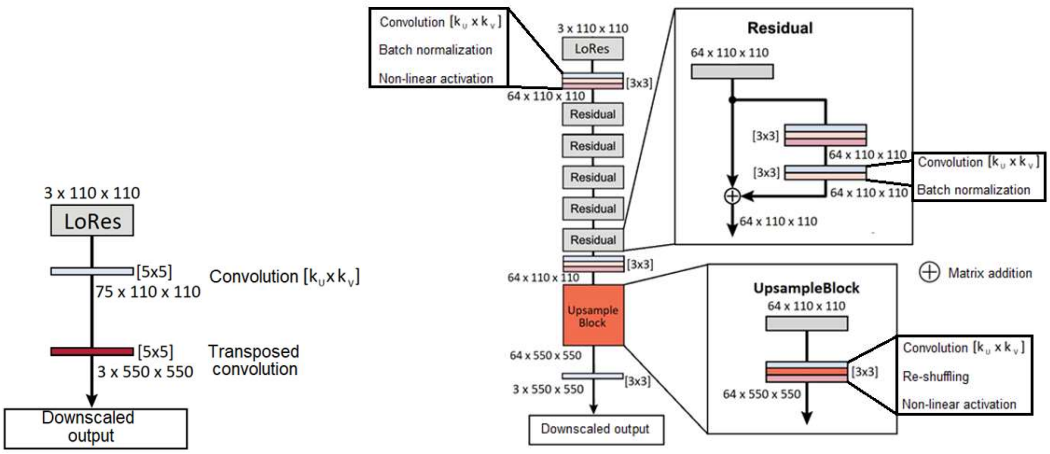
We chose convolutional neural network Linear CNN [1] as the simplest CNN in this work. Linear CNN links low-resolution input data to high-resolution output without non-linear activation functions. This architecture consists of one branch (Fig. 1a), and we can describe Linear CNN as an autoencoder. The first layer is an encoder and the second layer is a decoder converting the hidden representation into an explicit array of three-dimensional vectors.

Increasing depth of CNNs improves the quality of prediction and downscaling. However, such an increase leads to training instability of backpropagation algorithm. Learning becomes inefficient due to «vanishing gradients». This negative effect accumulates extremely small gradients of model parameters. As a result, the product of the gradient vector and the learning rate coefficient tends to zero, and the parameters updated at each optimization step remain constant. The effect of «vanishing gradients» remains a significant problem in ANNs.

A batch is a set of elements of the training dataset simultaneously processed by the model. After the model calculated the output data of a batch, the model parameters are updated, and the following batch is a next input for the model. Batch normalization is able to stabilize network learning and to correct output data distribution of the convolutional layer.

Another effective way to solve the problem of learning instability is to add connections that skip the intermediate layers of the model. Such connections reduce the possibility of small gradients accumulation. An example is a residual connection when the output of an intermediate layer is added to the output of a later level. Residual connections allow the model to learn outputs from initial layers at the beginning of the network.

We combined the advantages of deep CNNs, batch normalization and residual connections to study Residual CNN based on the EnhanceNet [1]. The proposed network included an input layer for extracting raw features, followed by a sequence of convolutional layers and residual blocks for processing the extracted features, and an upscaling block (Fig. 1b).



(a) (b)

Figure 1: (a) Linear CNN architecture; (b) Residual CNN architecture

Residual learning replaces part of the convolutional layers with residual blocks consisting of two branches. The first branch is an identical transfer of the data and the second branch is a sequence of two convolutional layers. The matrices obtained after both branches of the block are summed up. We also added parametric linear rectifier PReLU as nonlinear activation functions to approximate nonlinear relationships.

Transposed convolution retards learning, and leads to «checkerboard» effect in the model output. To solve these problems, we provided an upsampling block with convolutional layers and «pixel shuffle». This transformation returns the required spatial size of the matrix.

Another example of struggling with learning instability in deep neural networks is skip connections. In skip connections, two matrices of dimensions C1×W×H and C2×W×H are concatenated into a matrix of (C1+C2)×W×H dimension. The number of channels varies.

For wind downscaling, we consider surface wind systems as the result of a complex interaction between large-scale atmospheric dynamics and processes in the atmospheric boundary layer on smaller horizontal scales. Thus, the correct interpretation of physical processes for different horizontal scales is an important aspect of the correct wind downscaling.

Skip connections effectively extract functional dependencies corresponding to different spatial scales. We propose CNN architecture with different levels of abstraction to interpret data at several scales (Fig. 2).

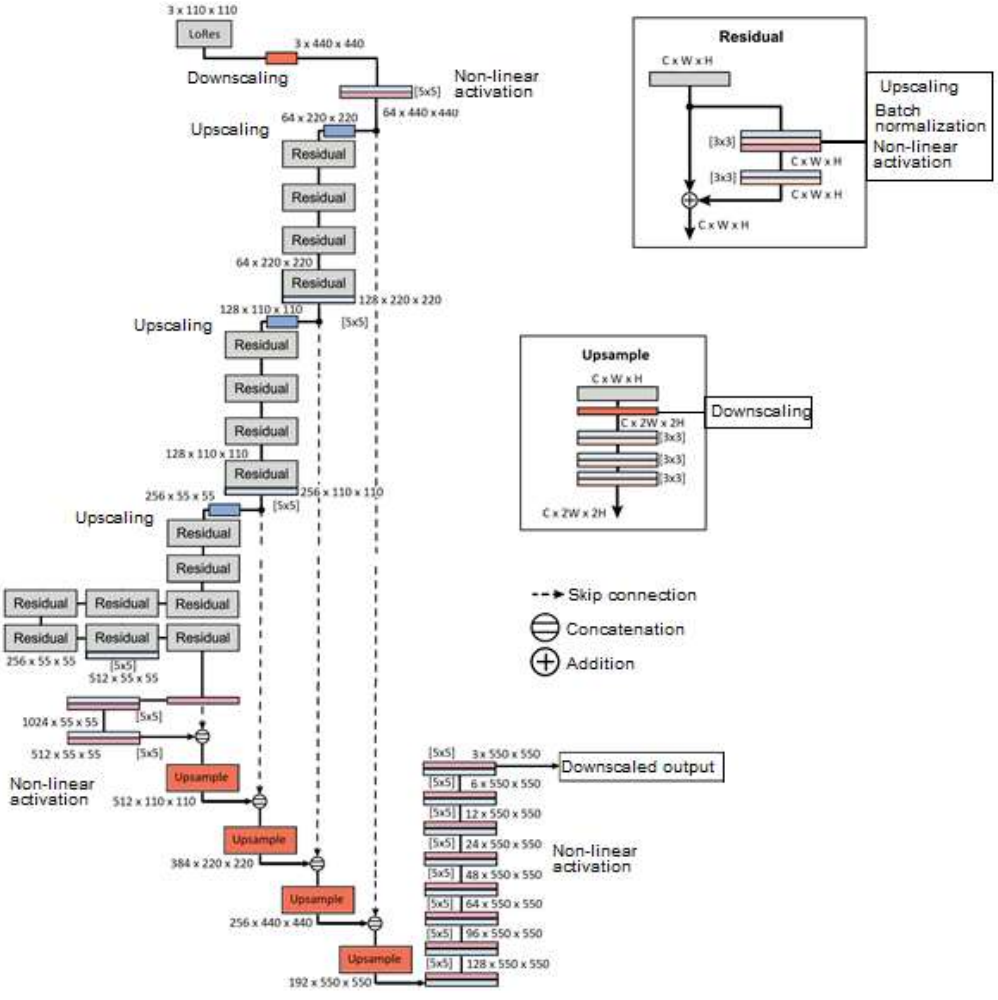


Figure 2: CNN with skip connections

The architecture of CNN with skip connections consists of two asymmetric branches and a convolutional block. The branches are linked by skip connections transferring information from the encoding branch to the decoding one. A large-scale data structure is extracted by small 5×5 convolution kernels. At the decoding stage, the spatial size of the matrices rises while simultaneously reducing the number of channels. Skip connections ensure the direct transfer of information between layers in the encoding and decoding branches with the same resolution.

4 Generative adversarial artificial neural networks (GANs)

Unlike discriminative ANNs, adversarial models are trained by an adversarial process where two networks are trained simultaneously. In the adversarial network architecture, the generator is opposed to the discriminator. The generator creates different outputs and learns to make them more plausible and similar to real output-like data. The discriminator learns to determine whether a certain element is taken from the distribution generated by the generator or from the true data distribution.

This concept uses different types of models and learning algorithms for the generator and discriminator. In particular, if generative and discriminative models are ANNs, the network as a whole is a generative adversarial network (GAN).

We studied the SR-GAN model [30] as an example of GAN-based downscaling. The architecture of the network we exploited is similar to the one presented in Figure 4 of the original paper [30] with the following correspondence: we use LoRes NAAD data snapshots as Input of the Generator network; HiRes NAAD snapshots correspond to HR reference data in SR-GAN; high-resolution statistically downscaled snapshots correspond to SR data in SR-GAN scheme [30].

The deep convolutional generator was similar to Residual CNN. The discriminator contained convolutional blocks, non-linear activation LeakyReLU (leaky rectified linear unit) and batch normalization. The resulting value of the discriminator was from 0 to 1. We interpreted the output value of the discriminator as the probability that the element was taken from the true data distribution.

5 Quality metrics

We considered various quality metrics to compare models in terms of downscaling quality.

We used the root mean-squared error (RMSE) of the wind speed as the simplest downscaling quality metric in this work:

$$RMSE = \sqrt{\frac{1}{550\times550}\sum_{i=1}^{550}\sum_{j=1}^{550}\left(\sqrt{\left(U_{i,j}^{*}\right)^2+\left(V_{i,j}^{*}\right)^2}-\sqrt{\left(U_{i,j}\right)^2+\left(V_{i,j}\right)^2}\right)^2},$$

where $U_{i,j}$ and $V_{i,j}$ are HiRes wind speed components, $U_{i,j}^{*}$ and $V_{i,j}^{*}$ are downscaled wind components on 550×550 grid.

RMSE showed standard deviation of the downscaled wind from the true values. The disadvantage of the RMSE quality metric was that it did not determine the quality of downscaling of strong winds usual for open-water regions.

We needed to resolve powerful near-surface flows over the ocean strongly localized in space. Because of these considerations, we introduced a new quality metric, RMSE-95. This metrics was

similar to RMSE but calculated errors only for wind values that were higher than 95-th percentile of the wind speed. Consequently, RMSE-95 showed standard deviation of the extremely strong downscaled wind.

Peak signal-to-noise ratio (PSNR) is another quality metric frequently used in image downscaling:

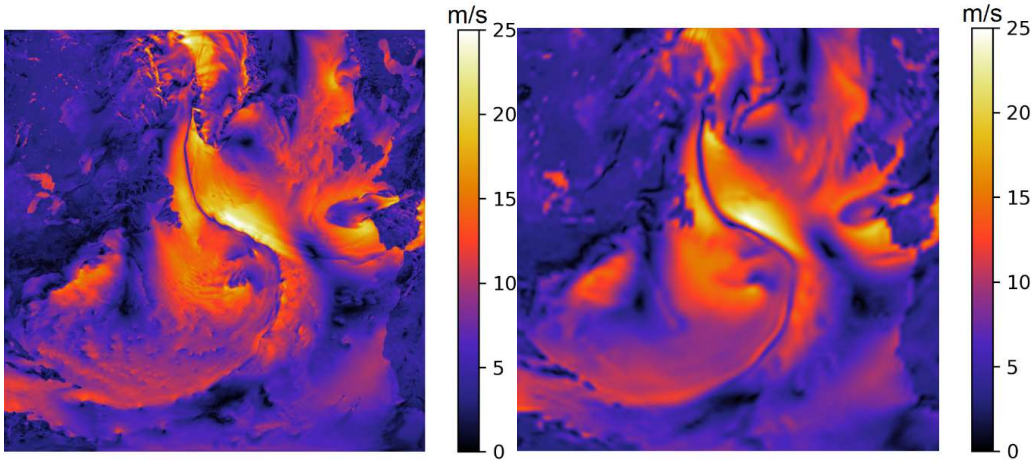
$$PSNR = 10 \log \left(\frac{MAX^2}{MSE} \right)^2,$$

where MAX is maximum normalized value between 3 downscaled variables on 550×550 grid, MSE is mean-square error of 3 normalized downscaled variables.

This metric relates to the perceived quality of the downscaled image. The value of the peak signal-to-noise ratio tends to infinity when the RMSE approaches zero. Higher PSNR value indicates a higher image quality. As the maximum value of the variables (signal) increases, the peak signal-to-noise ratio also increases. An increase in the signal, and, consequently, an increase in PSNR results in better visual perception of the downscaled climatic variables.

6 Downscaling results

Fig. 3a shows true HiRes wind W. The reference solution is bicubic interpolation W* (Fig. 3b). As an example, we select data at 00:00, 1 January 2010. This point of time is in the validation dataset. We choose the winter period due to stronger winds in the North Atlantic region to check model ability to resolve high-speed values.



(a) (b)

Figure 3: Wind speed (00:00, 1 Jan 2010), m/s:
(a) NAAD HiRes; (b) Bicubic interpolation

The difference (W*−W) defines RMSE quality metric. If this wind difference is calculated only for wind exceeding the 95-th percentile, it defines the RMSE-95 quality metric.

Fig. 3b shows that the bicubic interpolation is inaccurate. The interpolated wind greatly differs from its true values. The wind difference averaged over the entire region ⟨W*−W⟩ is close to zero (0.05 m/s) indicating that the interpolation equally smoothed both extremely low and extremely high values. The averaged difference $\langle W^* - W \rangle_{95}$ for extremely strong winds is −1.39. The negative sign proves a strong smoothing of extremely high values.

We present quality metrics for bicubic interpolation in Table 1.

	RMSE, m/s	RMSE-95, m/s	PSNR	$\langle W^*-W\rangle$, m/s	$\langle W^*-W\rangle_{95}$, m/s
00:00 1 January 2010	1.56	2.71	35.79	0.05	−1.39
Validation dataset	1.44	1.90	35.16	0.05	−0.55

Table 1: Bicubic interpolation quality

In general, quality metrics for the entire validation dataset show values that do not differ much from the specific data selected as an example. The relatively high values of RMSE-95 for the selected example are explained by the winter period and, as a result, stronger winds.

The simplest CNN studied in this paper is Linear CNN. Fig. 4a shows the same values for Linear CNN as Fig. 3b. Fig. 4a shows the noising «checkerboard» effect specific for linear architectures. The «checkerboard» effect is caused by transposed convolution increasing the image resolution. We move away from this resolution increase method in other CNNs studied in this work.

The downscaled wind in Linear CNN differs significantly from the NAAD HiRes wind. $\langle W^*-W\rangle$ averaged over the entire region is far from zero (−1.46 m/s), which differs from the bicubic interpolation. A bias towards smoothing high values of wind speed is specific for Linear CNN. This architecture greatly underestimates extremely strong winds: $\langle W^*-W\rangle_{95} = -5.69$.

We present the quality of the Linear CNN model in Table 2.

	RMSE, m/s	RMSE-95, m/s	PSNR	$\langle W^* - W\rangle$, m/s	$\langle W^* - W\rangle_{95}$, m/s
00:00 1 January 2010	3.30	7.35	35.93	−1.46	−5.69
Validation dataset	2.85	5.32	27.68	−0.79	−2.25

Table 2: Linear CNN quality

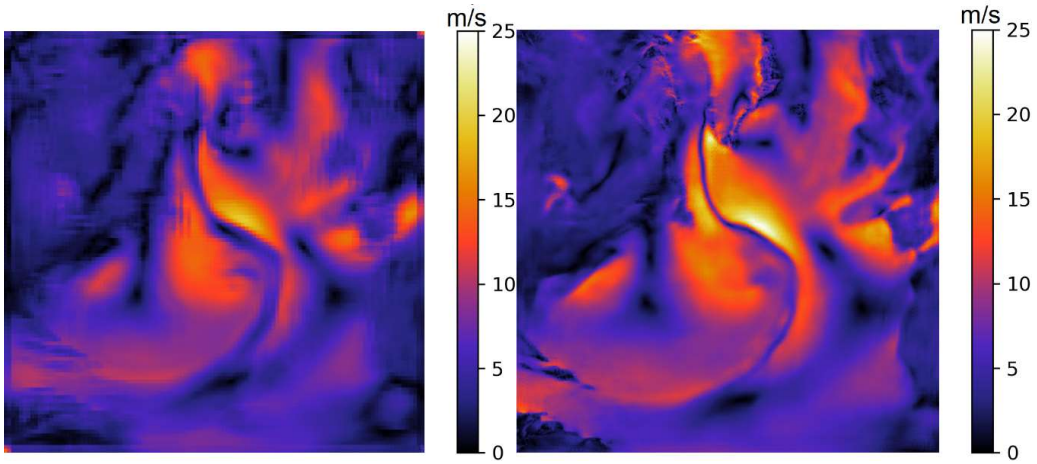
For the entire validation dataset, $\langle W^*-W\rangle$ is smaller in absolute value than for the chosen point of time. All four seasons are included and averaged in the dataset. In this case, the negative sign is important because it indicates the smoothing of extremely high values. In general, Linear CNN shows poorer results than bicubic interpolation.

The next model studied in this paper is Residual CNN. It has deeper architecture and residual connections. An example of Residual CNN downscaling (Fig. 4b) shows no «checkerboard» in downscaled wind. We explain it by the replacement of the transposed convolution with an upsampling block.

In general, the downscaled wind looks too smooth and does not resolve the fine-scale structure of the wind (Fig. 4b). The smoothing of extremely low and extremely high values is not as strong as in Linear CNN. In addition, Fig. 4b has fewer areas where the downscaled wind is close to the true one, and more areas with extremely large errors that confirm the downscaling quality improvement.

The Residual CNN model still underestimates wind values over the 95-th percentile, but the error of such values is smaller.

Residual CNN quality metrics are in Table 3. Comparison of Table 3 and Table 2 shows the advantage of Residual CNN over Linear CNN for all quality metrics.



(a) (b)

Figure 4: Wind speed (00:00, 1 Jan 2010), m/s:
(a) Linear CNN downscaling; (b) Residual CNN downscaling

	RMSE, m/s	RMSE-95, m/s	PSNR	$\langle W^* - W \rangle$, m/s	$\langle W^* - W \rangle_{95}$, m/s
00:00 1 January 2010	1.63	3.13	34.67	−0.39	−1.92
Validation dataset	1.42	2.21	32.87	−0.19	−0.25

Table 3: Residual CNN quality

The next complication of CNN in this paper is skip connections. CNN with skip connections contains residual blocks and batch normalization. Fig. 5a shows an example of CNN with skip connections downscaling.

Fig. 5a shows a significant downscaling improvement over the continents. We explain this result by more stable structure of the wind over land caused by orographic reasons. As a result, the model learns to properly downscale wind fields over continents rather than over water where the wind is stronger and subject to temporal variability. This implies the absence of fine-scale information in the downscaled wind fields over the Atlantic Ocean.

Fig. 5a shows that the downscaling error over land is smaller than over the Atlantic Ocean. CNN with skip connections underestimates the wind values at points where the wind exceeds the 95-th percentile.

The quality metrics of CNN with skip connections are in Table 4.

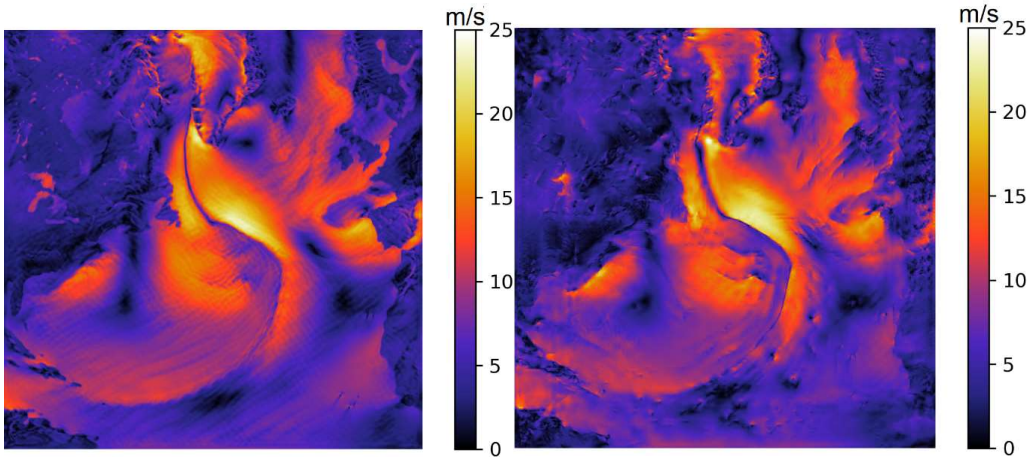
	RMSE, m/s	RMSE-95, m/s	PSNR	$\langle W^* - W \rangle$, m/s	$\langle W^* - W \rangle_{95}$, m/s
00:00 1 January 2010	1.40	2.43	36.44	−0,08	−1,41
Validation dataset	1.32	1.97	34.46	−0,02	−0,72

Table 4: CNN with skip connections quality

CNN with skip connections outperforms Residual CNN in all quality metrics. Better downscaling over the continents provides this. The only significant improvement is the value of

⟨W*−W⟩, which is closer to zero than all previous models are. This architecture smoothes very low and very high wind values equally, slightly changing the statistical distribution of wind speed in a modeling area.

Generative adversarial network (GAN) differs from previous discriminative CNN architectures because it consists of two models trained simultaneously in an adversarial process. GAN generator is based on Residual CNN and SR-GAN [30]. Fig. 5b shows an example of GAN generator downscaling.



(a) (b)

Figure 5: Wind speed (00:00, 1 Jan 2010), m/s:
(a) CNN with skip connections downscaling; (b) GAN generator downscaling

Fig. 5b shows the result different from previous discriminative models. Unlike discriminative CNNs, GAN generator allows getting a fine-scale structure of wind fields over the ocean. For example, the generator downscales the sequence of eddies in the south and southeast parts of the modeling area without smoothing. In addition, the structure in the center of the region is transferred correctly.

The largest GAN downscaling errors are over land. GAN discriminator prevents the model from overfitting typical for CNN with skip connections. On the contrary, GAN generator is trained so that it computes plausible high-resolution wind without smoothing the extreme values. Fine-scale downscaled structure is similar to the NAAD HiRes input, but pixel-by-pixel error is significant. That is an important disadvantage of GAN compared to discriminative models.

The quality metrics for GAN generator are shown in Table 5.

	RMSE, m/s	RMSE-95, m/s	PSNR	$\langle W^* - W \rangle$, m/s	$\langle W^* - W \rangle_{95}$, m/s
00:00 1 January 2010	2.40	4.73	37.59	−0.35	−3.28
Validation dataset	1.88	3.30	33.99	−0.22	−1.72

Table 5: GAN generator quality

GAN is worse than discriminative CNN methods studied in this work, in terms of quality metrics. However, GAN architecture is the most promising downscaling method. In this study, the GAN model is the only one detecting the fine-scale wind structure over the North Atlantic.

The summary of downscaling quality is in Table 6. The best values of every quality metric are bold.

Method	RMSE, m/s	RMSE-95, m/s	PSNR	$\langle W^* - W \rangle$, m/s	$\langle W^* - W \rangle_{95}$, m/s
Bicubic interpolation	1.44	**1.90**	**35.16**	0.05	–0.55
Linear CNN	2.85	5.32	27.68	−0.79	−2.25
Residual CNN	1.42	2.21	32.87	−0.19	**−0.25**
CNN with skip connections	**1.32**	1.97	34.46	**−0.02**	−0.72
GAN	1.88	3.30	33.99	−0.22	−1.72

Table 6: Downscaling quality (bold – best value of quality metric)

7 Discussion

This study does not suggest that neural-network downscaling methods will be used directly for operational prediction based on coarse-grid data. The results show that the resulting downscaling quality of discriminative and generative models is not competitive in terms of the spatial resolution compared to the existing dynamical methods. Nevertheless, GAN can give a promising basis for further development of statistical climate downscaling methods.

One of the problems to be solved is the required number of training predictors for a model. Datasets with a large number of climate variables and static predictors will allow the model to be trained more efficiently.

There are many applications of statistical downscaling techniques requiring more accurate local wind speed predictions. These include local distribution of air pollutants, sailing, etc. In cases where mean speed predictions are important, as for renewable energy, computationally cheap neural-network downscaling methods will be widely used. For precise values of extreme wind speeds, additional studies are required. For example, model training time can be insufficient, and downscaled forecasts will be smoothed.

Furthermore, including temporal information into the process of model building or model training could be an interesting direction for future research. The time-series coherence in predictions for given sites is necessary to be comparable to time-series coherence in the training data. The current time-independent approach is good in that it might preserve frontal passage wind-shifts at grid points, but on the other hand this may cause other unexplainable temporal shifts in wind speed.

We also find promising directions for further improvements of statistical downscaling approaches that imply additional regularizations restricting a network during training. In our study, no physics-based constraints are applied. At the same time, conservation laws can be injected into loss function. In addition, wind statistics or spectral characteristics can be used as regularization terms.

8 Conclusion

In this study, we analyzed artificial neural networks for downscaling of wind fields on extended North-Atlantic spatial domain. We went from a simple Linear CNN to deeper and more elaborate nonlinear discriminative models. After that, we analyzed generative adversarial model. We investigated how the network complexity affects downscaling performance and quality.

We demonstrated that deeper and more complex network models were not fully able to discover skillful wind mappings. We found that nonlinear discriminative models of our study demonstrated either no improvement or minor improvement compared to reference bicubic interpolation.

We strongly believe that the demonstrated performance of generative network for downscaling tasks should motivate further research towards the use of such non-classical architectures for predictive tasks.

In conclusion, we should note that the emergence of new methods for the problem of wind speed downscaling increases the scope of the forecasts obtained in this way, that is especially important for regions with complex topography. In turn, further research on neural-network methods will improve the quality by expanding their application in addition to numerical weather forecasting.

References

[1] K. Höhlein, M. Kern, T. Hewson, R. Westermann, *A comparative study of convolutional neural network models for wind field downscaling*, *Meteorological Applications* **27**(6) (2020) [https://doi.org/10.1002/met.1961]

[2] D. Maraun et al., *Towards process-informed bias correction of climate change simulations*, *Nature Climate Change* **7**(11) (2017), pp. 764–773 [https://doi.org/10.1038/nclimate3418].

[3] H. Hersbach et al., *The ERA5 global reanalysis*, *Quarterly Journal of the Royal Meteorological Society* **146**(730) (2020), pp.1999–2049 [https://doi.org/10.1002/qj.3803].

[4] B. C. Hewitson, R. G. Crane, *Climate downscaling: techniques and application*, *Climate Research* **7**(2) (1996), pp. 85–95 [https://doi.org/10.3354/cr007085].

[5] M. Rummukainen, *State-of-the-art with regional climate models*, *Wiley Interdisciplinary Reviews: Climate Change* **1**(1) (2010), pp. 82–96 [https://doi.org/10.1002/wcc.8].

[6] Y. Xue et al., *A review on regional dynamical downscaling in intraseasonal to seasonal simulation/prediction and major factors that affect downscaling ability*, *Atmospheric research* **147** (2014), pp. 68–85 [https://doi.org/10.1016/j.atmosres.2014.05.001].

[7] P. Ayar et al., *Intercomparison of statistical and dynamical downscaling models under the EURO- and MED-CORDEX initiative framework: present climate evaluations*, *Climate dynamics* **46**(3) (2016), pp. 1301–1329 [https://doi.org/10.1007/s00382-015-2647-5].

[8] A. Casanueva, S. Herrera, J. Fernández, J. M. Gutiérrez, *Towards a fair comparison of statistical and dynamical downscaling in the framework of the EURO-CORDEX initiative*, *Climatic Change* **137**(3) (2016), pp. 411–426 [https://doi.org/10.1007/s10584-016-1683-4].

[9] A. J. Cannon, C. Piani, S. Sippel, *Bias correction of climate model output for impact models*, *Climate Extremes and Their Implications for Impact and Risk Assessment,* Elsevier 2020.

[10] D. Maraun et al., *Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user*, *Reviews of geophysics* **48**(3) (2010), pp. 682–689 [https://doi.org/10.1029/2009RG000314].

[11] R. E. Chandler, *On the use of generalized linear models for interpreting climate variability*, *Environmetrics: The official journal of the International Environmetrics Society* **16**(7) (2005), pp. 699–715 [https://doi.org/10.1002/env.731].

[12] J. M. Gutiérrez et al., *Reassessing statistical downscaling techniques for their robust application under climate change conditions*, *Journal of Climate* **26**(1) (2013), pp. 171–188 [https://doi.org/10.1175/JCLI-D-11-00687.1].

[13] E. Hertig, J. Jacobeit, *A novel approach to statistical downscaling considering non-stationarities: application to daily precipitation in the Mediterranean area*, *Journal of Geophysical Research: Atmospheres* **118**(2) (2013), pp. 520–533 [https://doi.org/10.1002/jgrd.50112].

[14] D. San-Martín et al., *Reassessing model uncertainty for regional projections of precipitation with an ensemble of statistical downscaling methods*, *Journal of Climate* **30**(1) (2017), pp. 203–223 [https://doi.org/10.1175/JCLI-D-16-0366.1].

[15] A. W. Wood, L. R. Leung, V. Sridhar, D. P. Lettenmaier, *Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs*, *Climatic change* **62**(1) (2004), pp. 189–216 [https://doi.org/10.1023/B:CLIM.0000013685.99609.9e].

[16] E. Eccel et al., *Prediction of minimum temperatures in an alpine region by linear and non-linear post-processing of meteorological models*, *Nonlinear processes in geophysics* **14**(3) (2007), pp. 211–222 [https://doi.org/10.5194/npg-14-211-2007].

[17] C. F. Gaitan, W. W. Hsieh, A. J. Cannon, P. Gachon, *Evaluation of linear and non-linear downscaling methods in terms of daily variability and climate indices: surface temperature in southern Ontario and Quebec, Canada*, *Atmosphere-Ocean* **52**(3) (2014), pp. 211–221 [https://doi.org/10.1080/07055900.2013.857639].

[18] T. Vandal, E. Kodra, A. R. Ganguly, *Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation*, *Theoretical and Applied Climatology* **137**(1) (2019), pp. 557–570 [https://doi.org/10.1007/s00704-018-2613-3].

[19] D. A. Sachindra et al., *Statistical downscaling of precipitation using machine learning techniques*, *Nature Climate Change* **212** (2018), pp. 240–258 [https://doi.org/10.1016/j.atmosres.2018.05.022].

[20] A. Bhardwaj et al., *Downscaling future climate change projections over Puerto Rico using a non-hydrostatic atmospheric model*, *Climatic Change* **147**(1) (2018), pp. 133–147 [https://doi.org/10.1007/s10584-017-2130-x].

[21] J. Baño-Medina, R. Manzanas, J. M. Gutiérrez, *Configuration and intercomparison of deep learning neural models for statistical downscaling*, *Geoscientific Model Development* **13**(4) (2020), pp. 2109–2124 [https://doi.org/10.5194/gmd-13-2109-2020].

[22] H. Y. Huang, S. B. Capps, S. C. Huang, A. Hall, *Downscaling near-surface wind over complex terrain using a physically-based statistical modeling approach*, *Climate dynamics* **44**(1) (2015), pp. 529–542 [https://doi.org/10.1007/s00382-014-2137-1].

[23] P. A. Michelangeli, M. Vrac, H. Loukos, *Probabilistic downscaling approaches: Application to wind cumulative distribution functions*, *Geophysical Research Letters* **36**(11) (2009) [https://doi.org/10.1029/2009GL038401].

[24] S. C. Pryor, J. T. Schoof, R. J. Barthelmie, *Empirical downscaling of wind speed probability distributions*, *Journal of Geophysical Research: Atmospheres* **110** (2005) [https://doi.org/10.1029/2005JD005899].

[25] D. Maraun, M. Widmann, J. M. Gutiérrez, *Statistical downscaling skill under present climate conditions: A synthesis of the VALUE perfect predictor experiment*, *International Journal of Climatology* **39**(9) (2019), pp. 3692–3703 [https://doi.org/10.1002/joc.5877].

[26] D. J. Sailor, T. Hu, X. Li, J. N. Rosen, *A neural network approach to local downscaling of GCM output for assessing wind power implications of climate change*, *Renewable energy* **19**(3) (2000), pp. 359–378 [https://doi.org/10.1016/S0960-1481(99)00056-7].

[27] A. Gavrikov et al., *RAS-NAAD: 40-yr High-Resolution North Atlantic atmospheric hindcast for multipurpose applications (new dataset for the regional mesoscale studies in the atmosphere and the ocean)*, *Journal of Applied Meteorology and Climatology* **59**(5) (2020), pp. 793–817 [https://doi.org/10.1175/JAMC-D-19-0190.1].

[28] J. G. Powers et al., *The weather research and forecasting model: Overview, system efforts, and future directions*, *Bulletin of the American Meteorological Society* **98**(8) (2017), pp. 1717–1737 [https://doi.org/10.1175/BAMS-D-15-00308.1].

[29] D. P. Dee et al., *The ERA-Interim reanalysis: Configuration and performance of the data assimilation system*, *Quarterly Journal of the royal meteorological society* **137**(656) (2011), pp. 553–597 [https://doi.org/10.1002/qj.828].

[30] C. Ledig et al., *Photo-realistic single image super-resolution using a generative adversarial network*, *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 4681–4690 [https://doi.org/10.48550/arXiv.1609.04802].